Energy-Aware Scheduling at LRZ

EPAJSRM BoF@SC18 | 11/14/18 | Michael Ott

Motivation and History

- Electricity in Germany is expensive (~0.18€/kWh)
- Single budget for hardware procurement and energy for 6 years of operation
- Goal: saving energy without hurting performance

Energy-Aware Scheduling (EAS) on SuperMUC:

- Joint development with IBM for Load Leveler
- Based on user-provided "energy tags" to identify applications/work loads
- Characterization of application on first run based on HW performance counters
- Proxy applications to determine optimal frequency for lowest energy-to-solution

Future Plans

SuperMUC-NG
- Deployment in Q4/2018
- Intel Skylake, 6480 nodes, 96GB RAM, Omnipath
- 26.9PF
- #8 Top500 11/2018

- Retain and improve EAS capabilities
  - Run memory-bound codes at lower clock rates, compute-bound at maximum clock rate
  - Allow for multiple program regions with different characteristics
  - Avoid user interaction

- Deploy DataCenter DataBase (DCDB) system-wide high-frequency monitoring

- Use historical application runs to influence job scheduling:
  - Schedule "hot" jobs on adsorption chiller islands
  - Balance applications temporally and spatially